ARM's First Low-Power Superscalar Processor

David Williamson
Co-Architect and Validation Lead

Fall Processor Forum
October 2005

Market requirements for wireless

- Wireless applications are requiring more processing capabilities
  - Email, web, gaming, and cameras on everything
  - Video, image, speech, and music for multimedia
  - Graphics, physics, 2D, and 3D rendering for games

- Current processors are insufficient to meet these needs
  - Need more work done per unit time
    - Architectural efficiency and frequency both must improve
  - Power and energy are not free
  - Transistors are not free
    - Costs are rising for → 130nm, 90nm, 65nm, etc.

- There must be balance
  - Architecture, microarchitecture, and implementation must all be explored and utilized to the fullest

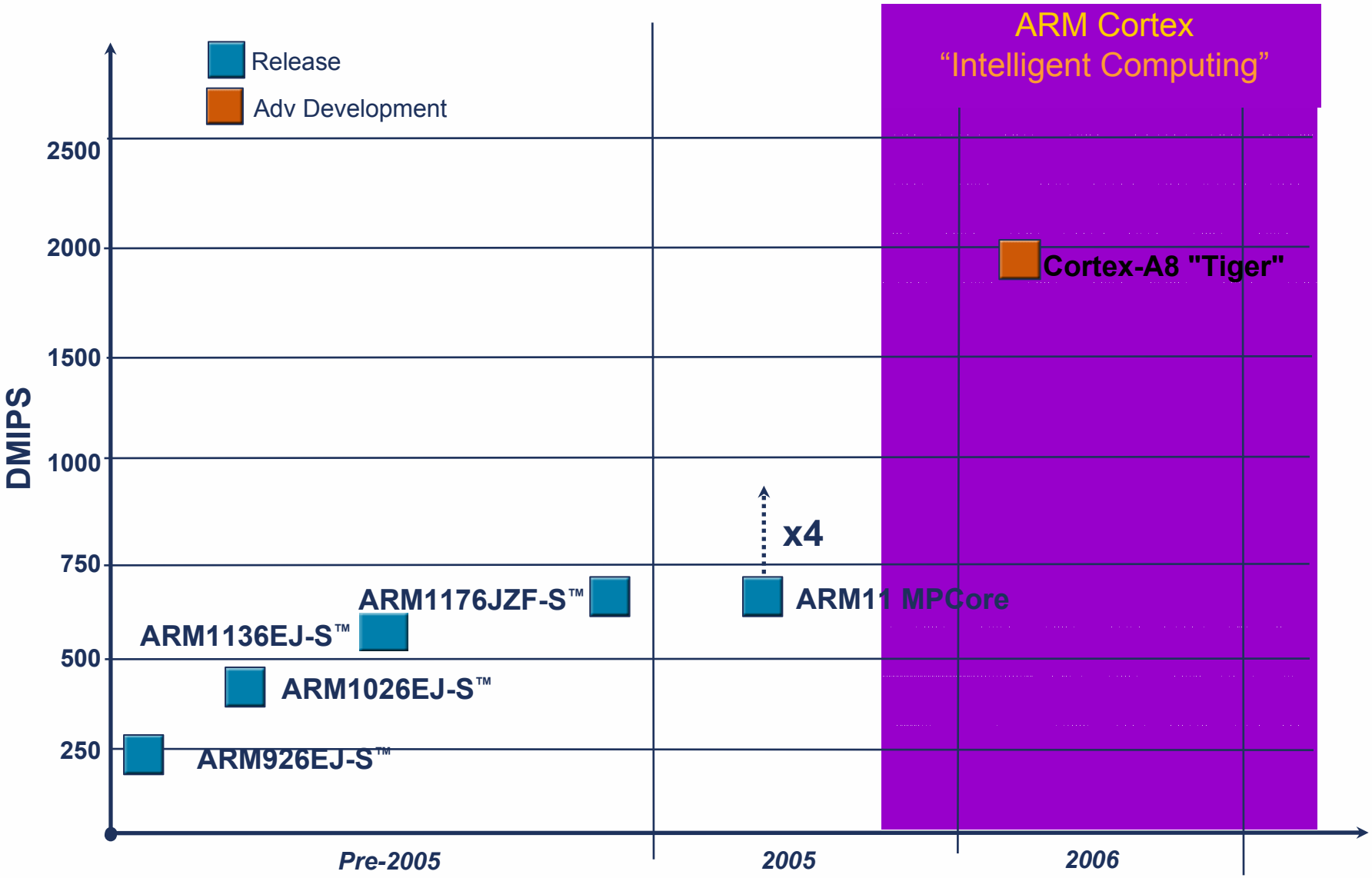
Cortex-A8 Processor Highlights

- First implementation of the ARMv7 instruction-set architecture, including the Advanced SIMD media instructions (NEON™)

- In-order, dual-issue, superscalar microprocessor core
  - 13-stage main integer pipeline
  - 10-stage NEON media pipeline
  - Dedicated L2 cache with 9-cycle latency
  - Branch prediction based on global history

- Key metrics
  - Delivers 2000 DMIPS for next-generation consumer applications
  - Average instructions per cycle (IPC) = 0.9 across multiple benchmark suites
    - Includes EEMBC, SPECint95, Mediabench, and partner-provided applications
  - Delivers 1GHz for consumer applications when fabricated in high-performance 90nm and 65nm process technologies
  - Consumes less than 300mW for mobile applications when fabricated in 65nm
  - Less than 4mm^2 at 65nm, excluding NEON, L2 cache, and Embedded Trace

ARM Applications Processor Roadmap



ARMv7 Architecture Overview

- Consolidates established components implemented in transition products
  - TrustZone®
  - Thumb®-2
- Jazelle-RCT
  - Efficient acceleration of execution environments using just-in-time (JIT) and dynamic adaptive compiler (DAC) technologies
  - Delivers JIT performance with code size comparable to original bytecodes
  - Minimal hardware cost — fewer than ten new instructions, reuse Thumb-2 decoders
- Advanced SIMD Media Architecture (NEON)
  - Accelerates audio, video, imaging, graphics, and speech algorithms
  - 64-bit SIMD data processing
  - Integer and floating-point data types
  - Provides 2x–4x performance improvement over ARMv6 SIMD

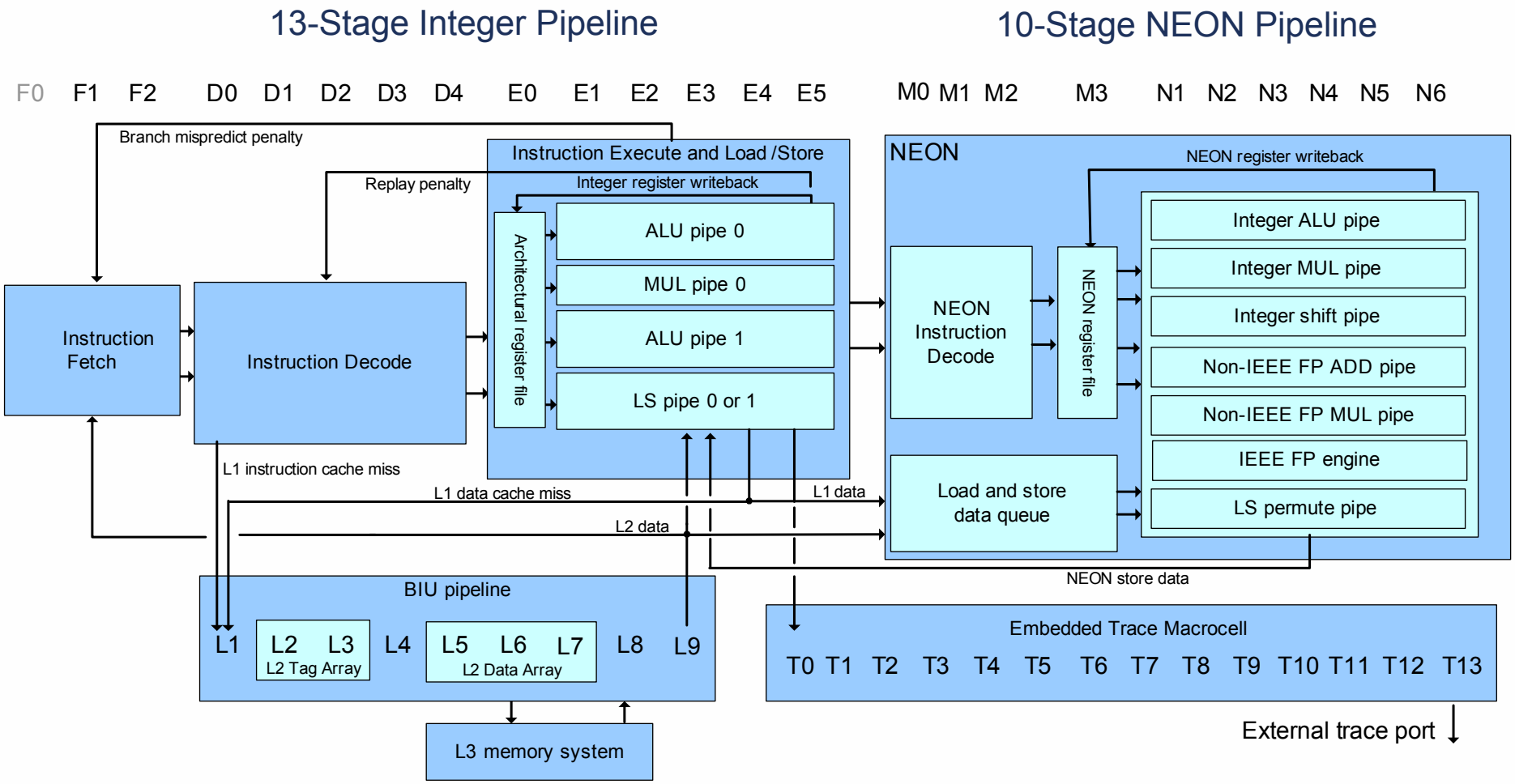
Superscalar design tradeoffs

- In-order instruction issue
  - Less complex than out-of-order
    - Fewer structures means lower power
    - Less need for custom design
  - Can still maintain high IPC with
    - Fully symmetric ALU pipelines
    - All critical forwarding paths supported
    - Dual-issue of dependent instruction pairs

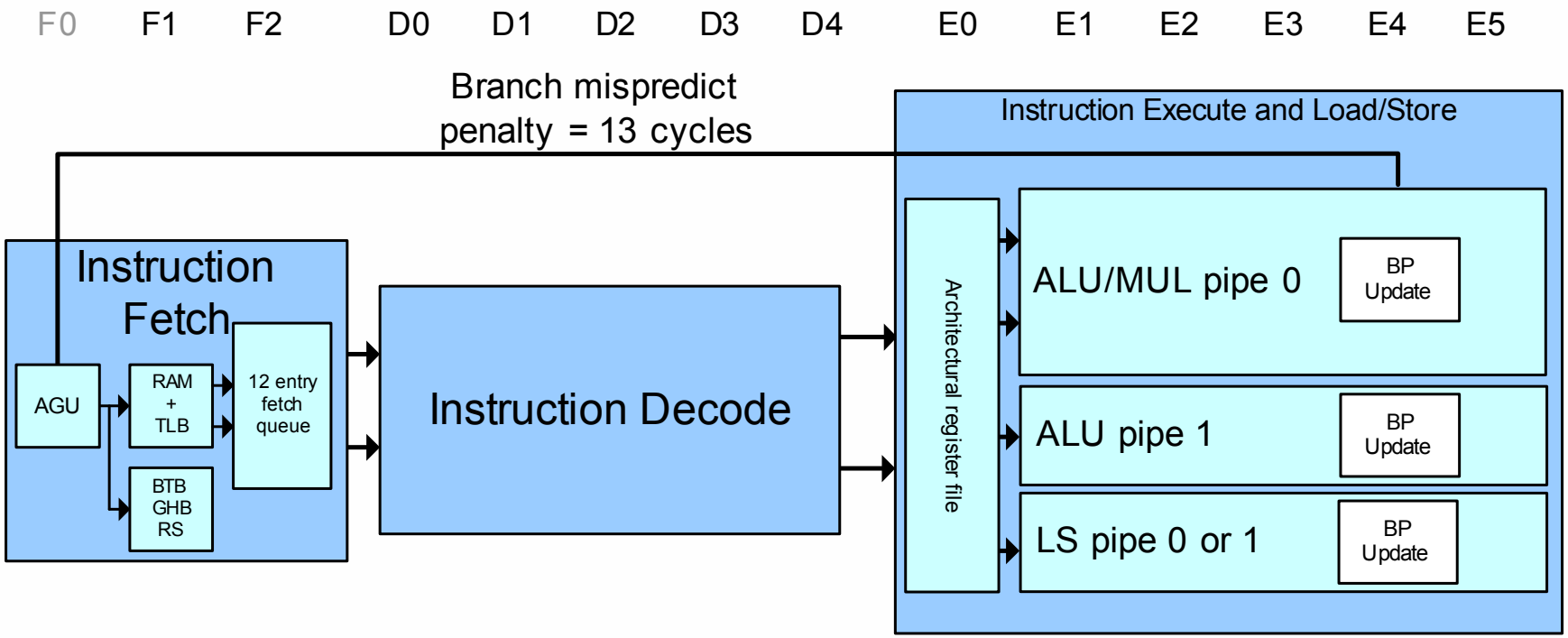
- Static scheduling with instruction replay on memory stall
  - Low power consumption due to early availability of gate enables
  - Fire-and-forget instruction issue removes key critical paths from the design

- Net result
  - *High-frequency design with out-of-order performance, but in-order clock frequency and power consumption*

Full Cortex-A8 Pipeline Diagram



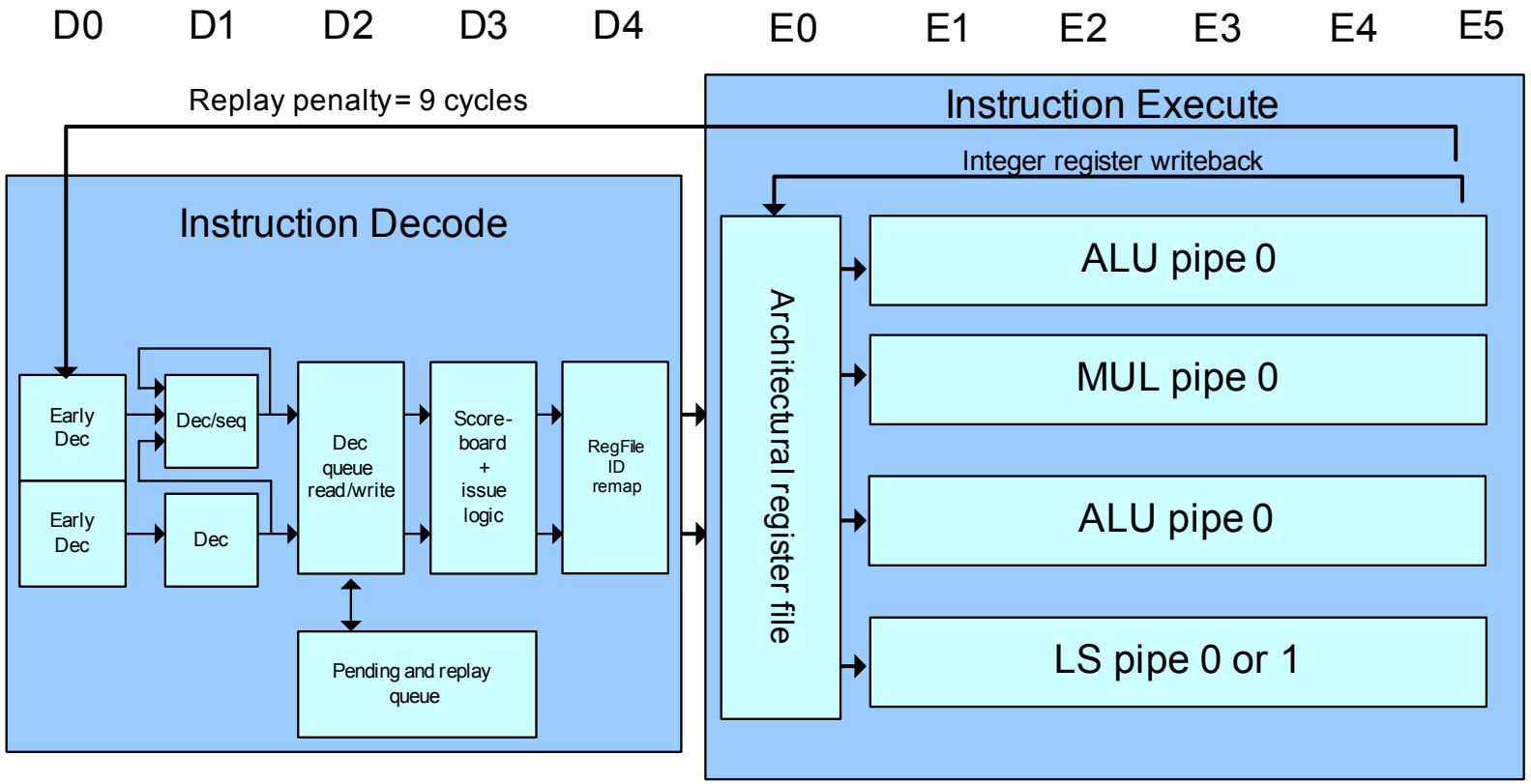
Control Flow



- Dynamic branch predictor components
  - 512-entry 2-way branch target buffer (BTB)
  - 4K-entry GHB indexed by branch history and PC
  - 8-entry return stack
- Branch resolution
  - All branches are resolved in single stage
  - Maintains speculative and nonspeculative versions of branch history and return stack

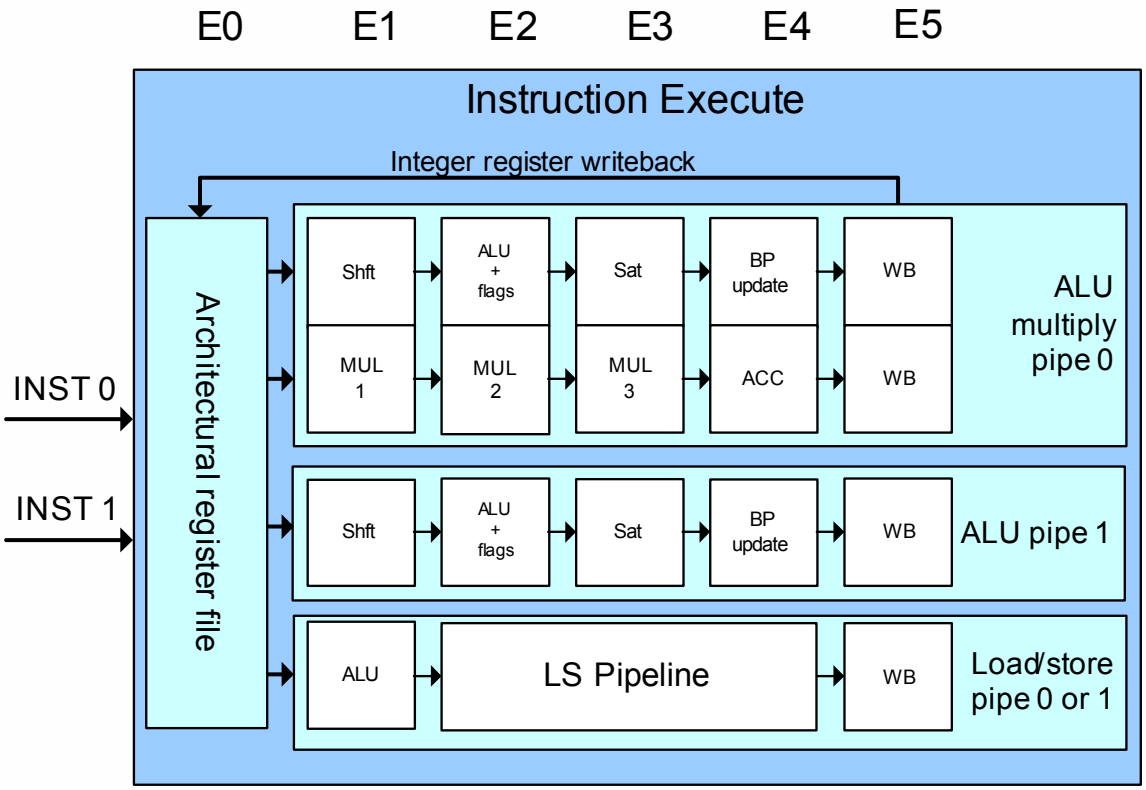
Branch prediction maintains 95% accuracy over a wide code base

Instruction Decode



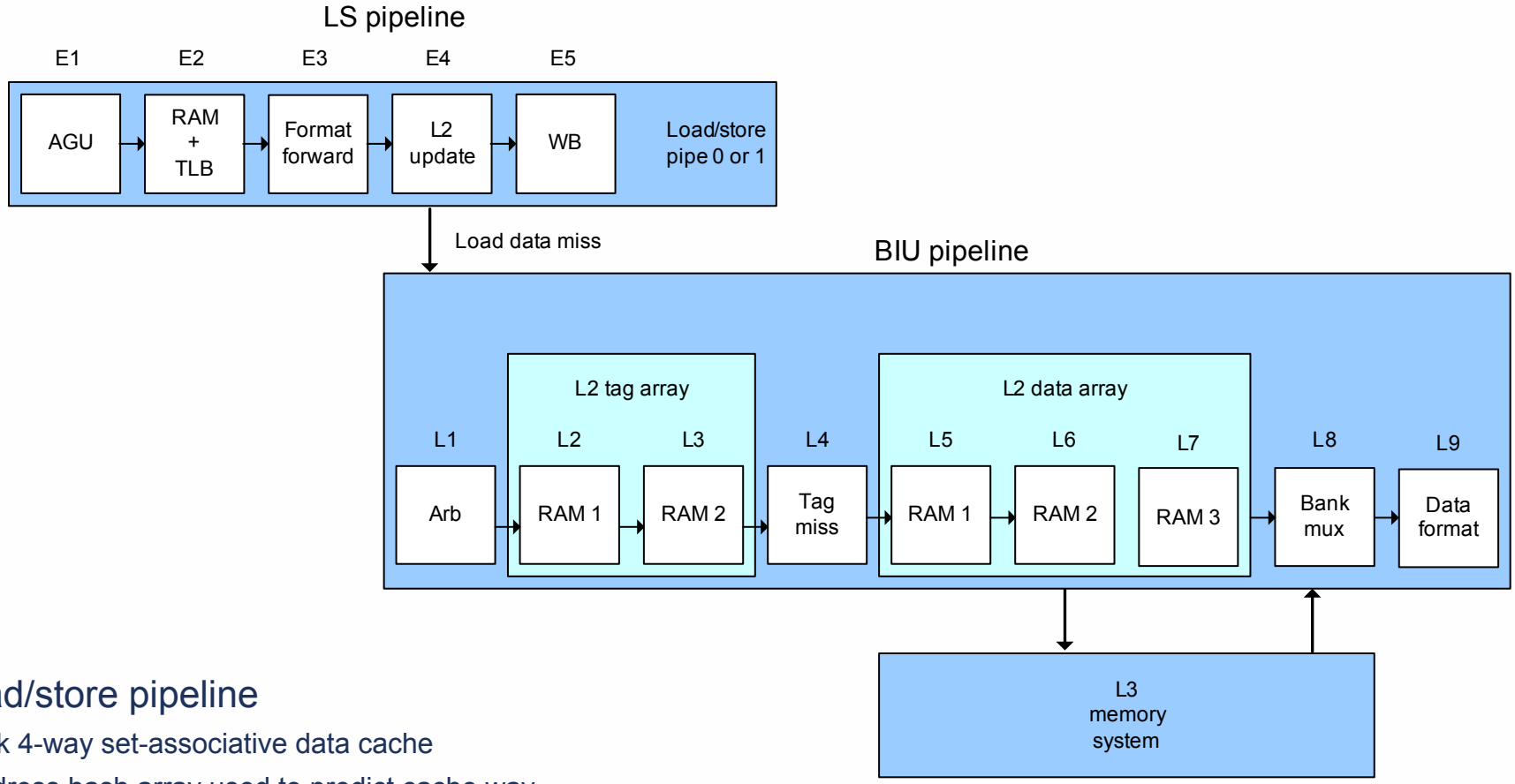
- Instruction decode highlights
  - Pending queue reduces fetch stalls and increases pairing opportunities
  - Replay queue keeps instructions for reissue on memory system stall
  - Scoreboard predicts register availability using static scheduling techniques
  - Cross-checks in D3 allow issue of dependent instruction pairs

Instruction Execution



- Execution pipeline highlights
  - Two symmetric ALU pipelines: Shift/ALU/SAT
  - Load/store pipe used by instructions in either pipeline
  - Multiply instructions are tied to pipe 0
  - All key forwarding paths supported
  - Static scheduling allows for extensive clock gating

Memory System



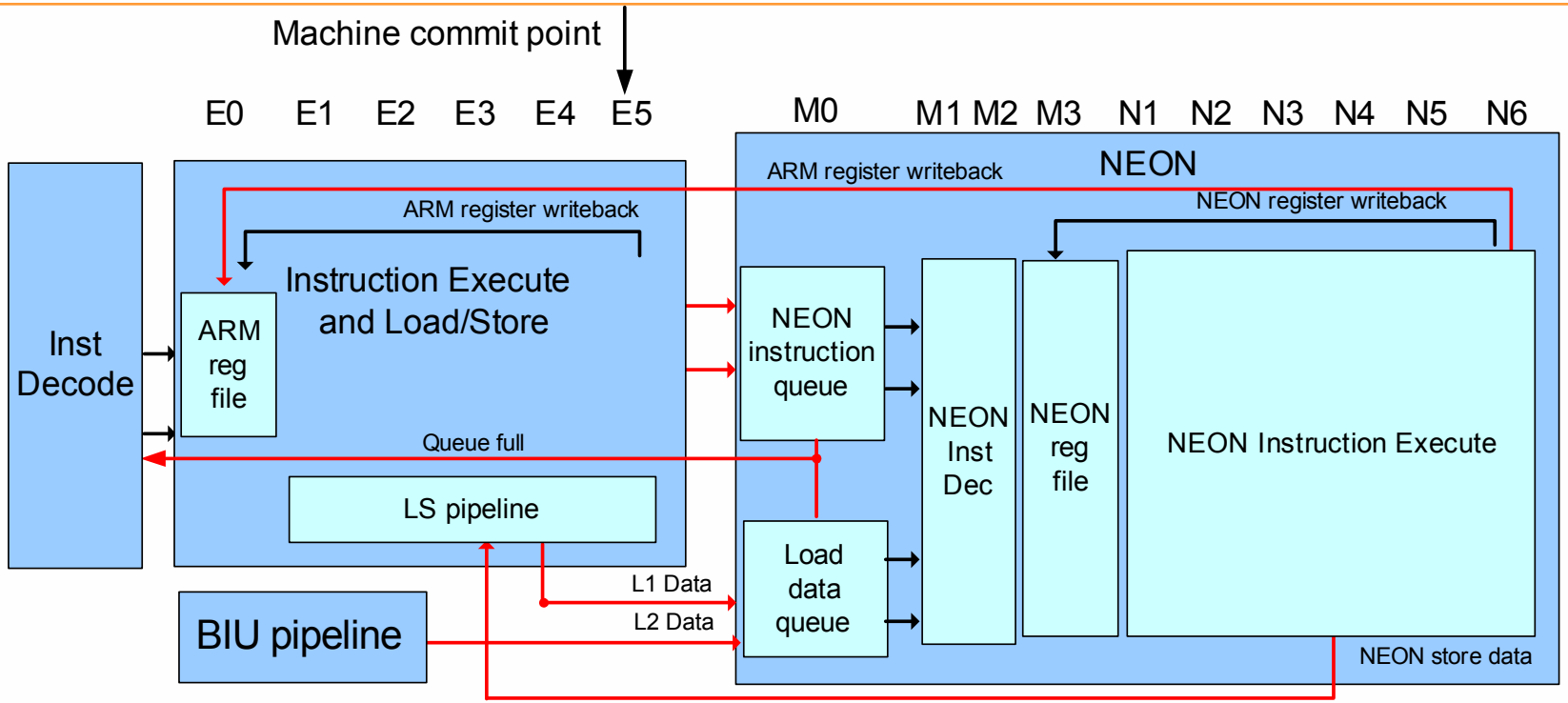
Load/store pipeline

- 32k 4-way set-associative data cache
- Address hash array used to predict cache way
  - Saves power and improves timing
- Load data forwarding in E3 to all critical sources
  - One-cycle load-use penalty for ALU
- Store data not required until E3

BIU pipeline

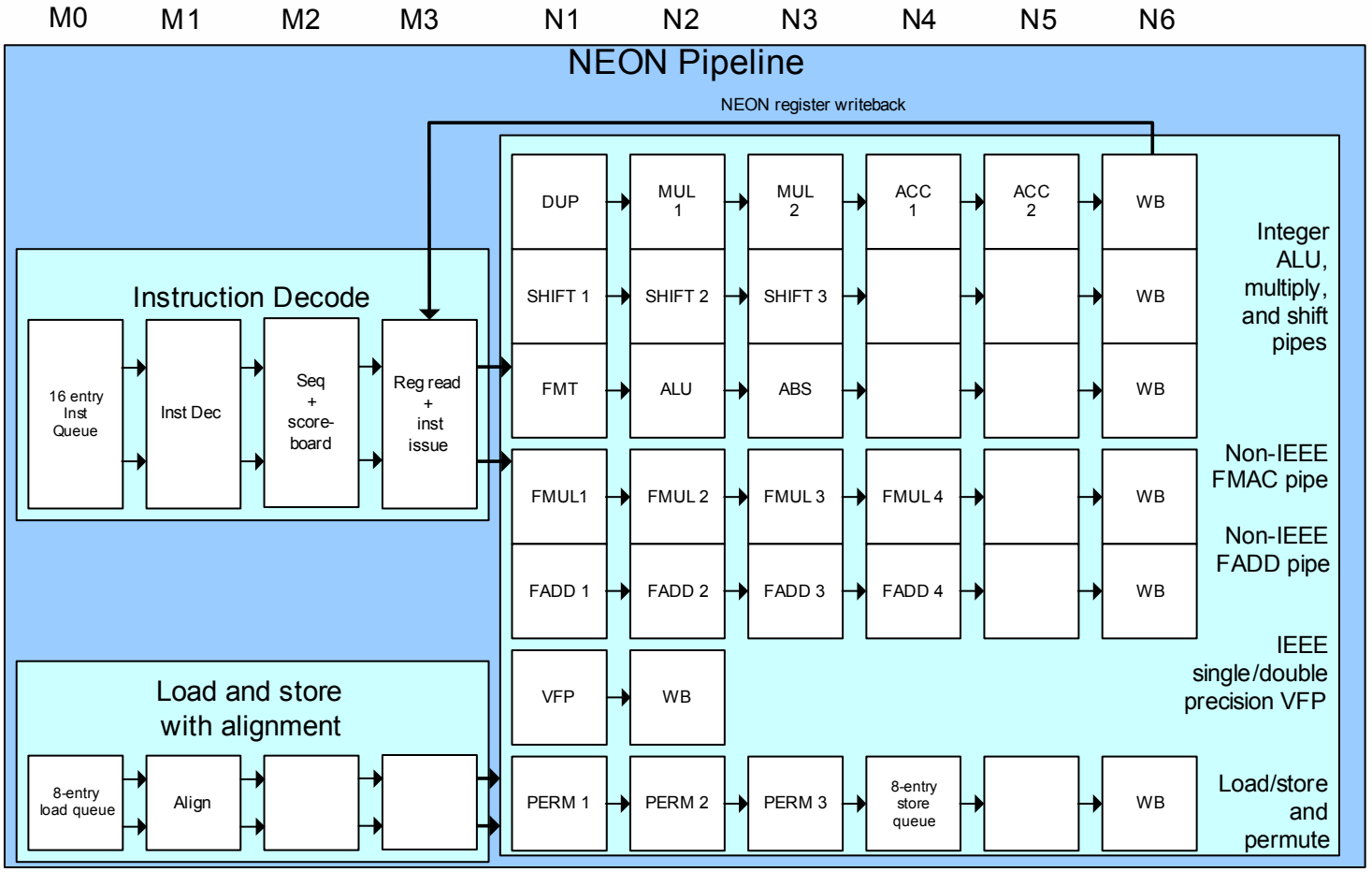
- 9-cycle minimum access latency to L2 cache
- L2 built using standard compiled RAMs (64k–2MB configurable size)
- 64/128bit AXI L3 bus interface supports up to 9 outstanding transactions

NEON Interfaces



- Skewed late in pipeline, past the retire point
  - Reduces interface complexity
    - Exception handling not required
    - Decoupling queues from integer machine
  - Removes load-use penalty
  - Negative impact on NEON -> ARM transfers
    - Nonblocking ARM register file helps hide latency
- Streaming to and from L2 memory system
  - Up to 8 outstanding transactions
  - Can receive 128 bits/cycle
  - Can receive data from L1 or L2 memory system
  - Independent NEON store buffer

NEON Media Engine Unit



Instruction issue

- Static scheduling with fire-and-forget issue
- 1 LS + 1 NINT/NFP can issue each cycle

Execution pipelines

- All pipelines are 64-bit SIMD
- Floating-point MAC executed using both FADD and FMUL pipelines

Implementation Strategy: Motivation

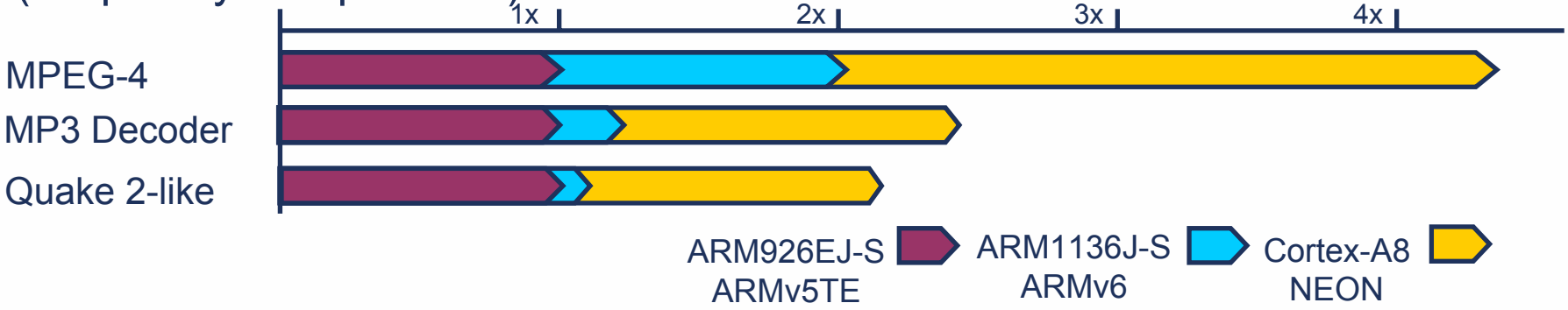
- Why use a semicustom design flow?
  - Required to achieve combined improvements in power, performance, and area
- Why not deliver a hard macrocell?
  - Fixed-process hard macrocell has too many restrictions on product flexibility
  - Process-portable design does not scale well with increases in design size and complexity
- The goal:
  - Provide our partners with an alternative method of IP delivery that
    - Achieves Cortex-A8 power, area, and frequency targets
    - Minimizes the additional effort required from the silicon partner

Implementation Strategy

- Microarchitecture tuned for high-performance implementation
  - 10 gates (NAND3 equivalent) allowed per cycle
  - Synthesized, structured, and custom circuits all utilized to balance design
  - A nontraditional cell library, AdvantageCE, to support both synthesized and structured cell-based design
- Implementation flow balances ease of development with high-performance targets
  - Cell-based flow and methodology used to speed development
  - Industry standard tools and flows used wherever possible
  - Specialized IP deliverables provided
    - State-accurate, synthesizable RTL model for formal equivalence
    - Gate-level reference netlist
    - Detailed specifications for custom circuits and cell library
- 9 customs total in design that focus on critical functions

Cortex-A8 NEON Performance

- Cortex-A8 NEON performance vs. ARMv5 and ARMv6 implementations (frequency independent)



- CPU bandwidth required for various applications:
  - MPEG4 VGA decode[1] 107 MHz
  - MPEG4 VGA decode with filters and color conversion[2] 296 MHz
  - MP3 decode, 320kbps 48kHz, worst case[3] 10 MHz
  - "Quake 2-like" application, CIF resolution[4] 304 MHz

1) MPEG-4 Simple Profile @ 30fps 512kbps , 133MHz SDRAM 10-1-1-1

2) MPEG-4 Simple Profile as (1), but also including dering and deblock filters and yuv2rgb conversion

3) MP3 Decoder @ 320kbps 48kHz (worst case means cold start on context switch), 133MHz SDRAM 10-1-1-1

4) Quake2-like simulator, full software graphics pipeline, floating point implementation, 30fps, 133MHz SDRAM 10-1-1-1

ARM Cortex-A8 Processor Summary

- First implementation supporting ARMv7 with NEON and Jazelle-RCT
  - Supports established ARMv6 technologies: TrustZone, Thumb-2, AXI, IEM
- Most efficient ARM processor
  - Superscalar core delivering an average IPC = 0.9 and 2.0 DMIPS/MHz
- Highest performing ARM embedded core
  - Achieves 1GHz with high-performance process technologies
- Delivers industry-leading power efficiency
  - Consumes less than 300mW in low-power mobile devices
- Widely supported by ARM technology
- Full release to lead partners 1H06

Cortex
Intelligent Processors by ARM

Artisan
ARM® Physical IP

RealView
ARM® Tools by ARM